REPORT PROJECT

Implementation Levenshtein Distance Algorithm to Searching Word

Brigitta Petrastuti
06.02.0020
2011

FAKULTAS ILMU KOMPUTER
UNIVERSITAS KATOLIK SOEGIJAPRANATA
Jl. Pawiyatan Luhur IV/1, Bendan Duwur, SEMARANG 50234
Telp. 204-8441555 (hunting) Web: http://www.unika.ac.id/
Email: ikom@unika.ac.id

APPROVAL AND RATIFICATION PAGE

PROJECT REPORT
Implementation Levenshtein Distance Algorithm to Searching Word

This Project Report has been approved and ratified by Dean of Computer Science Faculty on
….....

With the approval,

Examiner,

Rosita Herawati ST, M.IT
NIP: 058.1.2004.263

Examiner,

Robertus Aji Setiawan, ST, McompIT
NIP: 058.1.2004.264

Examiner,

Hironimus Marlon Leong , S.Kom, M.Kom
NIP: 058.1..2007..273

Supervisor,

Gregorius Hendita AK,
NIP: 058.1.2008.277

Dean of
FacultyConputer Science

Hironimus Marlon Leong , S.Kom, M.Kom
NIP: 058.1..2007..273

STATEMENT of ORIGINALY

The undersigned:

Name : Brigitta Petrastuti

ID : 06.02.0020

Here by certify that this project was made by my self and not copy or plagiarizes from other people, except that in writing expressed to the other article.

If it proven that this project was plagiarizes or copy the other, I'm ready to accept a saction.

Semarang, January 15th, 2011

Brigitta Petrastuti

06.02.0020

FOREWORD

I dedicated this graduation to my whole family that had been support me.

I would like to say thank you to:

- to God, that always guide me at making this project,
- to my parents, who always pray for me and support me to finish my project,
- to Gregorius Hendita A.K, as my supervisor who always helping, guiding and give me the ideas to finish this project,
- to my friends that always help me to support and helping me.

Last, I want to apologize if I've made some mistake at finishing this project and writing this respon. Therefore, critics and suggestions are expected.

Semarang, January 15th, 2011

Brigitta Petrastuti

ABSTRACTION

Levenshtein distance is a measure of the similarity between two strings, which we will refer to as the source string (s) and the target string (t). The distance is the number of deletions, insertions, or substitutions required to transform s into t. The greater the Levenshtein distance, the more different the strings are.

Levenshtein distance is named after the Russian scientist, Vladimir Levenshtein, who devised the algorithm in 1965. If you can't spell or pronounce Levenshtein, the metric is also sometimes called edit distance. The Levenshtein distance algorithm has been used in spell checking, speech recognition, DNA analysis, plagiarism detection .

This project is used for searching word from one file and after that this program will counting the word that was found, so users in addition to knowing whether the word exists or not, user will know too how much word in file whose user inputed. User only push the open button or user can choose the menu and click open in menu file, and choose the file that user want to search. After then, user can fill the keyword in text field, and then push the search button. So, the program will be printed result, the word is exist or not exist in the file, and if exist the result can be shown too how many word that exist in that file.

This project, is made using java language and levenshtein distance algorithm. However, this project in rough and still own feebleness which need corrected.

For example, if string (s) is "test" and string (t) is "test", then levenshtein distance of string(s,t) is 0, because no transformations are needed. The strings are already identical.

Another example if string (s) is "test" and string (t) is "tent", then

levenshtein distance if string(s,t) is 1, because one substitution (change "s" to "n") is sufficient to transform string (s) into string (t).

Keyword: searching word, levenshtein distance algorithm

TABLE OF CONTENT

TABLE OF FIGURES

TABLE OF TABLES